

Automatic identification of genes driving microevolution in evolution experiments

Scientists in the lab often perform so-called “evolution experiments”, in which they put a microbe in a condition it is not used to and wait until it adapts to the condition on a genetic level (evolution). Because this “evolution” occurs very rapidly and does not lead to new species but only to a small number of adjustments in the function of the microbe, this is called microevolution.

After the evolution experiment has been completed, data is generated by sequencing the genome from the evolved microbe and collecting information about the expression of genes in the microbe (RNA-seq or microarray). This leads to huge amounts of data which are cumbersome to interpret by hand. That is why algorithms which automatically analyze these big datasets are needed to guide further research. An important question such algorithms have to answer is: which of the (sometimes many) observed mutations are causal for the adaptation to the new environment? In order to provide an adequate answer to this questions we developed an algorithm called “PheNetic”.

In order to answer this question PheNetic uses the “genome-wide interaction network” of the microbe under research. This is a blueprint of the microbe which contains information on how its genes interact with each other. The interaction network is needed because normally an evolution experiment is repeated several times. Each time, the outcome (phenotype) of the organism will be very similar but the underlying mutation(s) causing this outcome might not be identical. While not identical, these mutations tend to occur in the same biological pathway and are thus close to each other in the genome-wide interaction network. PheNetic exploits this property by finding relevant “paths” in the genome-wide interaction network which overlap with each other.

Figure 1 gives an overview of PheNetic: it uses the expression data of genes to reweigh the interaction network, yielding a probabilistic subnetwork in which every interaction between two genes has a “chance” to be involved. Then, it searches for paths on the interaction network between mutated genes and highly up-or-down regulated genes. It does these steps for every repeat of the evolution experiment (focal end points). In a final step, using a probabilistic framework, PheNetic integrates all possible paths in all focal end points and selects a subnetwork containing the most probable paths with the highest overlap between all focal end points.

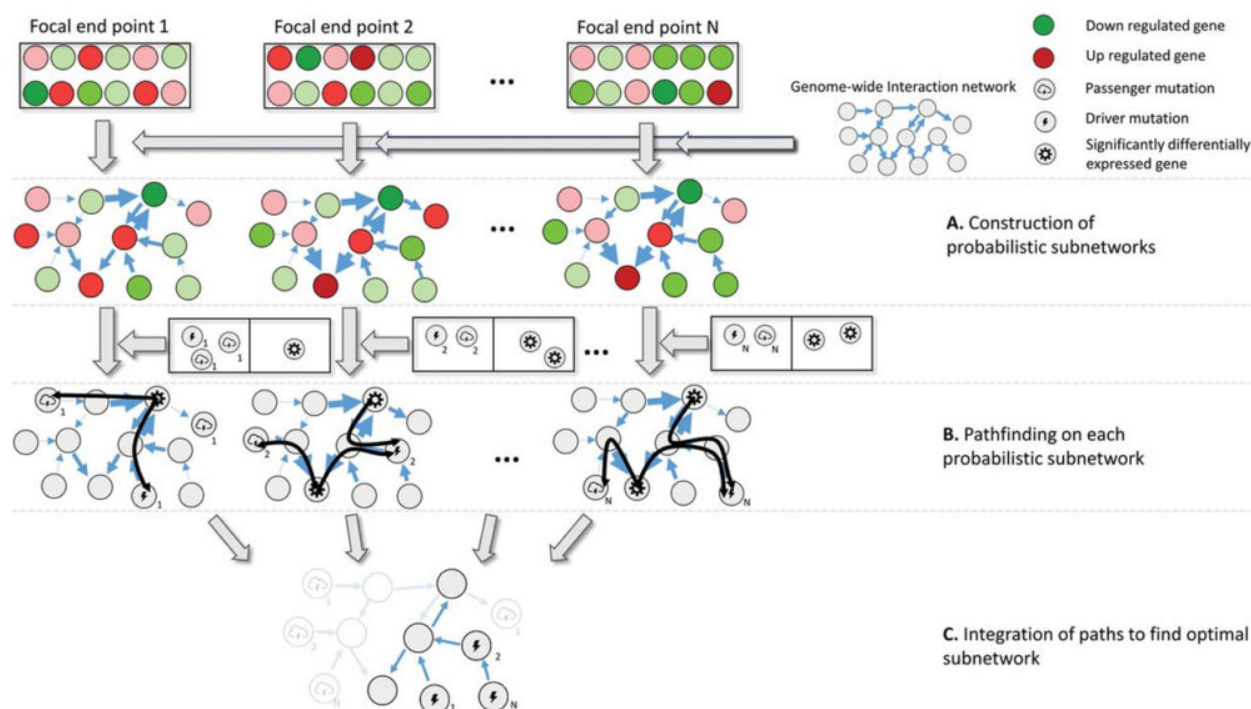


Fig. 1. PheNetic overview

We analyzed two very different evolution experiments using PheNetic: acquiring of antibiotic resistance (Amikacin) in *Escherichia coli* and the evolution of a population of *Escherichia coli* into a population containing two morphologically and functional distinct “ecotypes” of *Escherichia coli* which could coexist. In the case of Amikacin resistance we were able to automatically identify several genes involved in NADH dehydrogenase and terminal chain oxidase, both processes that can influence the proton-motif force, on which Amikacin uptake depends. In the coexisting ecotypes case, it was already known that two specific genes contributed to the co-existence (*arcA*, a component of the tricarboxylic acid cycle and *SpoT*, a component of the stringent response). PheNetic identified both genes as being of importance for the PheNotype and interestingly also identified another gene (*acs*) which is an extracellular acetate scavenger implying acetate as partly responsible for the cross-feeding mechanisms between the two ecotypes.

In conclusion algorithms such as “PheNetic” are important to make feasible and speed up the analysis of big data. This is especially important when large-scale experiments are conducted in order to find important genetic mechanisms which exhibit a low signal because e.g. they are rare. Important example which could be analyzed using this technology in the future are large-scale cancer datasets (thousands of patients) to prioritize drug targets and ethanol resistance datasets to improve the production of bioethanol.

Bram Weytjens, Dries De Maeyer, Kathleen Marchal

Department of Information Technology (INTEC, iMINDS), UGent, 9052 Ghent, Belgium

Department of Plant Biotechnology and Bioinformatics, Ghent University,

Technologiepark 927, 9052 Gent, Belgium

Bioinformatics Institute Ghent, Technologiepark 927, 9052 Ghent, Belgium

Department of Microbial and Molecular Systems, KU Leuven,

Kasteelpark Arenberg 20, B-3001 Leuven, Belgium

Publication

[Network-Based Analysis of eQTL Data to Prioritize Driver Mutations.](#)

De Maeyer D, Weytjens B, De Raedt L, Marchal K

Genome Biol Evol. 2016 Jan